ECONOMIC RESEARCH REPORTS

PROPORTIONAL REPRESENTATION: BROADENING THE OPTIONS

by **Steven Brams**
and
Richard Potthoff

RR# 97-06 **January 1997**

C.V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, NY 10003-6687

PROPORTIONAL REPRESENTATION: BROADENING THE OPTIONS

Steven J. Brams
Department of Politics
New York University
New York, NY 10003

Richard F. Potthoff
Center for Demographic Studies
Duke University
Durham, NC 27708

Steven J. Brams gratefully acknowledges the support of the C. V. Starr Center for Applied Economics at New York University.

PROPORTIONAL REPRESENTATION: BROADENING THE OPTIONS

Abstract

Using Monroe (1995) as a springboard, we extend and generalize his system of proportional representation (PR) by developing a general method for determining a set of winners from the ballots. Central to our analysis is the use of integer programming, which is a type of linear programming. Under Monroe's system and our generalizations of it, one minimizes total misrepresentation, where misrepresentation is based on approval votes, the rankings of candidates, or other ballot information. Our method allows for a variety of PR systems, including those proposed by Monroe, Chamberlin and Courant (1983), and Tullock (1967), as well as a new system we call "hierarchical PR." Ties, the filling of vacancies, and certain problems of both large and small electorates are all rendered manageable with integer programming. We discuss nonmanipulability, representativeness, and other criteria for selecting a PR system and conclude with some recommendations.

JEL Classification: D72. *Keywords:* Proportional representation; voting; approval voting; manipulability; integer programming.

In a provocative article entitled "Fully Proportional Representation," Monroe (1995) proposed a system of proportional representation (PR) that selects the winning candidates in an election for a legislature by minimizing the sum of "misrepresentation" values of all voters. These values are based on information provided by the voters on their ballots—information such as rankings of the candidates, approval votes, or other types of assessments.

Monroe's idea is a novel and intriguing one, but he left many stones unturned and a few misplaced. For example, he offered no mathematically explicit and verifiable method of achieving the intended minimization. Without such a method, there is a danger of selecting a set of winners that fails to minimize total misrepresentation; in fact, this deficiency plagued some of Monroe's own calculations. In addition, there are several ways, not covered by Monroe, in which his system can be extended so as to generalize the notion of representation he posited.

We begin by showing how integer programming provides a mathematical procedure for selecting winners so as to minimize the sum of misrepresentation values. Then we use integer programming to extend Monroe's system in several directions. Thus, instead of allowing just one candidate to be assigned to each voter in determining the solution of the minimization problem, we allow any number k up to the total number of candidates to be elected. We also examine, in more detail than did Monroe, the ramifications of using approval voting to achieve PR. Our PR systems include one with special properties that involves the use of approval voting in 3-member districts.

We analyze the effects of Monroe's system both in large electorates, where computational streamlining may be desirable, and in small electorates, where the chance of ties is the greatest. We consider the possibility of fractional vote assignments that force all winning candidates to have exactly the same number of voters assigned to them—rather than numbers that, in some cases, can differ from one another by 1, as allowed by Monroe.

The integer-programming approach can also handle restrictions *less* stringent than those imposed by Monroe concerning how nearly they equalize the numbers of voters assigned to winning candidates. At the extreme, one might invoke *no* restrictions, in which case Monroe's system becomes the same as that of Chamberlin and Courant (1983), which Monroe discusses (and rejects). If, further, one modifies the Chamberlin-Courant system by imposing no limit on the number of candidates to be elected, the system becomes the same as that of Tullock (1967, chap. 10), which is mentioned in the Chamberlin-Courant article but not in the Monroe article. Both the Chamberlin-Courant system and the Tullock system, however, are open to the criticism that they use weighted voting, a matter that we shall return to later.

Although integer programming is, as we will show, an extremely useful technique for implementing different PR systems, which system is best is by no means obvious. We will consider the pros and cons of both weighted-voting and nonweighted-voting PR systems. In addition, we will analyze the degree of nonmanipulability and the degree of representativeness of PR voting systems, both of which are important criteria in the choice of a voting system. Finally, we propose a new system of "hierarchical PR" that offers voters representation at different levels (e.g., president, senator, representative)—but it requires only one election to do so.

It would be derelict of us not to raise the practical question of whether PR voting systems of the kind considered here stand any chance of being used in real elections. Clearly, several obstacles exist. A major one is the inertia of overcoming the status quo, which is probably less of an obstacle for small electorates than for large ones, and for new legislative bodies than for established ones. Other obstacles include the complexity and the computational problems of PR systems; the latter will surely be ameliorated as computer capabilities continue to advance. Finally, educating voters about PR systems, and their possible advantages in certain situations, will be no small task, especially in the United States, where PR is quite unfamiliar.

There is one obstacle to the implementation of any PR system that has not been generally recognized. Ever since *Baker* v. *Carr* (1962) and related apportionment cases, the division of a jurisdiction into single-member districts has required that each district have almost the same total population, where population means people of all ages. Although one might question whether districting should instead be based only on total adult population, or perhaps even on total number of registered voters or actual voters, the current requirement is well established.

By contrast, PR representation in legislatures is based on voters rather than on population. A switch from single-member districts to PR is likely to lead to reduced voting strength in legislative bodies for ethnic and other groups that have low ratios of (i) adult to total population, (ii) registered voters to all adults, or (iii) actual voters to registered voters. Of course, there might be offsetting aspects of PR that could operate in such a way as to increase the voting strength of an affected group, but undoubtedly some groups will view PR, based as it is on actual votes cast, as likely to undermine their legislative strength—and hence will resist it as a threat to their positions.[1]

INTEGER PROGRAMMING

Because an integer program is a special type of linear program, we begin by defining the latter. In *linear programming*, one wants to find nonnegative values of the t variables $x_1, x_2, ..., x_t$ that maximize or minimize the linear function

$$z = c_1x_1 + c_2x_2 + ... + c_tx_t$$

subject to a set of s linear constraints that can be written as

$$a_{h1}x_1 + a_{h2}x_2 + ... + a_{ht}x_t \leq, =, \text{ or } \geq b_h \quad (h = 1, 2, ..., s).$$

The function z is called the *objective function.*

Integer programming, or *integer linear programming*, is a standard tool in certain disciplines such as management science (e.g., Davis and McKeown 1984, chap. 8; Gould, Eppen, and Schmidt 1993, chap. 9) but is probably unfamiliar to most political scientists.[2] An integer program is simply a linear program with an additional set of constraints: the t x-variables must all be integers. A *zero-one integer program* is an integer program in which each x-variable is constrained to be either 0 or 1. In a *mixed-integer program*, or *mixed-integer linear program*, some but not all of the x-variables are constrained to be integers.

Computer software packages are available to solve integer programming problems.

Although the mathematical techniques used by the packages are complicated, one does not generally need to understand these techniques in order to use the software and obtain the solutions.

Before showing how integer programming can be applied to PR problems, we introduce it with a simple non-PR example that illustrates the general kind of problem it is intended to solve. A merchant advertises bags of emeralds for sale at a certain price per bag, with the stipulation that each bag will contain no more than 4 emeralds totaling at least 65 units in weight. A customer arrives at a time when the merchant's inventory consists of 7 emeralds weighing 8, 8, 12, 15, 22, 32, and 41 units. Which emeralds does the merchant include in the bag for this customer so that the advertised conditions are observed yet the total weight is as low as possible?

The problem can be set up and solved as a zero-one integer program. There are $t = 7$ x-variables to be solved for, where x_i ($i = 1, 2, ..., 7$) is 1 if the i-th emerald is to be included in the bag and 0 if it is to be left out. The objective function, here to be minimized, is

$$z = 8x_1 + 8x_2 + 12x_3 + 15x_4 + 22x_5 + 32x_6 + 41x_7.$$

Other than the requirements for each x_i to be either 0 or 1, there are only two constraints:

$$8x_1 + 8x_2 + 12x_3 + 15x_4 + 22x_5 + 32x_6 + 41x_7 \geq 65,$$

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 \leq 4.$$

(Although the first constraint here has the same coefficients as the objective function, this will not generally be the case in an integer program.) The solution turns out to be $x_3 = x_5 = x_6 = 1$ and $x_1 = x_2 = x_4 = x_7 = 0$. That is, the emeralds selected are those with weights of 12, 22, and 32. Their total weight, $z = 66$, is not below 65; and the number of emeralds, 3, does not exceed 4.

MONROE'S SYSTEM

Let there be M candidates, m of whom are to be elected, and n voters. We use i ($i = 1, 2, ..., M$) to index the candidates and j ($j = 1, 2, ..., n^*$) to index the voters, where for present purposes $n^* = n$. For later purposes (in Appendix A), n^* can be less than n when we provide for j to index not just single voters but *groups* of voters who all vote the same way. Let there be a misrepresentation value, denoted by μ_{ij}, that reflects the extent to which voter j would be misrepresented by candidate i. As Monroe notes, if each voter is asked to rank the candidates from 1 (most preferred) to M (least preferred),[3] then μ_{ij} can be 1 less than the rank applied by voter j to candidate i; thus, if voter 5 ranks candidate 4 in third place, then $\mu_{45} = 2$. If approval voting (Brams and Fishburn 1983) is used, then μ_{ij} can be 0 if voter j approves of candidate i, 1 otherwise.

For any set of m out of the M candidates, Monroe assigns one of these m candidates to each voter in such a way that the number of voters linked to each candidate is n/m, or a number that differs from n/m by less than 1 if n/m is not an integer. If a unique set of m candidates produces an assignment that minimizes the sum of the misrepresentation values across all voters, it is declared to be the set of m winners. Although this criterion is well defined, Monroe did not provide a mathematically proven way to find the m winners so as to minimize misrepresentation in all cases.[4]

Setting up the problem as an integer program does provide such a way, which we now demonstrate. Let x_i be 1 if candidate i is a winner, 0 otherwise. Let x_{ij} be 1 if candidate i is assigned to voter j, 0 otherwise. Note that the x-variables are now of two types: M of them have one subscript, and Mn^* have two. Let the objective function be

$$z = \Sigma_i \Sigma_j \mu_{ij} x_{ij},$$

which is the sum of the misrepresentation values. It is to be minimized subject to the constraints

$$\Sigma_i x_i = m, \quad (1)$$

$$\Sigma_i x_{ij} = 1 \text{ for each } j \ (j = 1, 2, ..., n^*), \quad (2)$$

$$-Lx_i + \Sigma_j x_{ij} \geq 0 \text{ for each } i \ (i = 1, 2, ..., M), \quad (3)$$

$$-Ux_i + \Sigma_j x_{ij} \leq 0 \text{ for each } i, \quad (4)$$

$$x_i \text{ is an integer less than or equal to 1 for each } i, \quad (5)$$

$$x_{ij} \text{ is an integer less than or equal to 1 for each } (i, j) \text{ combination.} \quad (6)$$

Constraint (1) simply says that there are to be m winners out of the M candidates.

Constraint (2) specifies that one candidate is to be assigned to each voter.

Constraints (3) and (4) say that the number of voters to whom a candidate is assigned is to be greater than or equal to L but less than or equal to U for a winning candidate ($x_i = 1$); for a losing candidate ($x_i = 0$), (4) forces this number to be 0. If n/m is not an integer, then L is set equal to the largest integer less than n/m, and U to the smallest integer greater than n/m. If n/m is an integer, then (3)-(4) can still be used, with $L = U = n/m$, or (3)-(4) can be replaced by the single set of constraints

$$-Bx_i + \Sigma_j x_{ij} = 0 \text{ for each } i, \tag{7}$$

where $B = n/m$.

There are ways to create the misrepresentation values (the μ_{ij}'s) besides basing them on rankings or on approval votes. For example, each voter could rate each candidate on a $(T + 1)$-point scale from 0 to T, with μ_{ij} set equal to T minus the rating. Approval voting is a special case of such a system, with $T = 1$.

Another possibility would be to give each voter a fixed number of votes to spread across the candidates in any way desired, including giving more than one vote to the same candidate. Then μ_{ij} could be the negative of the number of votes that voter j gives to candidate i, or perhaps some other decreasing function of this number. The voting rules would be like those for cumulative voting, but the rules for determining the winners would not, because the latter rules would be based on misrepresentation rather than on total votes. We explored this type of system, but we encountered complexities and discovered nothing useful. Cumulative voting itself, however, has recently been used in several small jurisdictions in the United States to provide for minority representation.

For $m = 1, 2, 3$, and 4 winners, Monroe (1995) applied his voting system to an election in which the $M = 6$ candidates were actually newspapers; they were ranked by $n = 33$ voters. We ran integer programs in SAS® to solve for the winning sets of candidates for up to 4 winners (following Monroe) as well as for $m = 5$ winners. Our solutions appear in the first column of

Table 1A, which shows the winning candidates along with the associated total misrepresentation values. (The remaining columns in Table 1A, as well as all of Table 1B, will be explained shortly.) Our results do not fully agree with those of Monroe, as previously indicated in note 4. We believe integer programming is the only general technique that ensures that one always obtains a correct answer.

ASSIGNING MORE THAN ONE CANDIDATE TO A VOTER

It appears reasonable to generalize Monroe's system so as to allow k candidates ($1 \leq k \leq m$), rather than just one, to be assigned to each voter for purposes of determining the set of winners that minimizes total misrepresentation. Integer programming can easily accommodate such a generalization. One simply replaces relation (2) with

$$\Sigma_i x_{ij} = k \text{ for each } j. \tag{8}$$

Then in (3)-(4) L will now be the largest integer less than or equal to kn/m, and U will be the smallest integer greater than or equal to kn/m. If kn/m is an integer, (3)-(4) can be replaced by (7) with $B = kn/m$.

A value of k greater than 1 may be useful with approval voting, as we will see shortly. But for approval voting as well as other voting systems, one must guard against making k too high. If k is equal to its maximal value of m, then by (7) each winner is assigned to all n voters, making the winning candidates simply those whose μ_{ij} totals across all voters are lowest. (Two examples follow.) As a consequence, minorities, even large ones, may receive underrepresentation, even zero representation, if a disciplined majority votes in ways that crowd out the opposition. Thus, one may want to avoid $k = m$ to ensure minority representation.

As an example, consider an at-large election for an m-member local legislative body, where each voter is allowed to vote for up to m candidates. Our integer-programming framework applies, with $k = m$; define μ_{ij} to be 0 if voter j votes for candidate i, 1 otherwise. The integer program need not actually be run, because the winners will simply be the m candidates with the most votes. Thus, a 51% majority, all voting for the same m candidates, will win all the seats, however the 49% minority votes.

If the μ_{ij}'s are based on rankings and $k = m$, Monroe's system picks the m winning candidates to be those with the m highest Borda scores. Monroe (1995, 928) noted this result for the case $k = m = 1$.

For Monroe's voting data based on rankings, we ran integer programs with the constraints of (1), (3)-(6), and (8), for all values of m from 1 to 5 and all values of k from 1 to m. The results appear in Table 1A. For a given m, the sets of winners in a few cases change slightly as k increases, but mostly they stay the same.

To explain how integer programming produced the results in Table 1A, consider the cell for $m = 4$ and $k = 2$. It shows F, G, I, and T to be the four winners: the integer-programming solution yielded $x_2 = x_3 = x_5 = x_6 = 1$ and $x_1 = x_4 = 0$, where the subscripts $i = 1, 2, 3, 4, 5$, and 6 refer, respectively, to candidates D, F, G, H, I, and T. The solution for the 198 x_{ij}'s assigns the candidate pairs (F, G), (F, I), (F, T), (G, I), (G, T), and (I, T), respectively, to 5, 3, 8, 8, 3, and 6 voters. Thus, F is assigned to 16, G to 16, I to 17, and T to 17 voters. The total misrepresentation, 70 units, is the result of summing the 66 μ_{ij}'s that correspond to each voter's assigned candidates. For example, voter 3, who is assigned to candidates G and I, ranked the candidates in the order (I, T, G) (with no preference shown among the other three candidates) and thereby brought about values of (4, 4, 2, 4, 0, 1) for μ_{13} through μ_{63}. Hence, the contribution of voter 3 to the total misrepresentation is 2 + 0, or 2.

Note that total misrepresentation in this case has to be at least 33, because each voter will contribute at least a value of 1. More generally, total misrepresentation when rankings are used with Monroe's PR system cannot be less than $nk(k - 1)/2$.

APPROVAL VOTING

For possible use with his system, Monroe gave only limited consideration to approval voting but was optimistic about its prospects. We have made a more extensive study, and our conclusions are mixed.

First, we tried out approval voting using Monroe's data. His data do not provide a totally realistic test, because the voters were asked to furnish rankings rather than approval votes. One therefore has to make some assumption about how the voters, based on their rankings of candidates, would have voted had they used approval voting.

Assume that the p top-ranked candidates on each ballot, or each of the ranked candidates if fewer than p were ranked, would have received an approval vote. Because casting approval votes for about half the candidates is often a good strategy (Brams and Fishburn 1983, chap. 5), we started by setting $p = 3$ (recall that $M = 6$ newspapers were ranked). But we quickly encountered a problem, one that is likelier to arise the higher m, the lower k, or the higher p is: several sets of m candidates all had the same total misrepresentation—namely, zero. Thus, for $k = 1$ and $p = 3$ in Monroe's data, 9 of the 15 possible sets of $m = 4$ candidates and 5 of the 6 possible sets of $m = 5$ candidates produced total misrepresentation of zero, reflecting the fact that an approved candidate could be assigned to every voter. This embarrassment of riches—too many sets tied for winning—is hardly acceptable.

The apparent remedies are either (i) to lower p or (ii) to raise k. To be sure, in a real-world election using approval voting, one cannot reduce p, in the sense that one cannot alter the ballots. On the other hand, voters faced with a high m or a low k might well realize that their ballots will have a better chance of being decisive if they approve of fewer candidates, which corresponds to reducing p.

In any case, we used Monroe's system with approval voting, letting p vary according to m and k. The integer-programming solutions appear in Table 1B, which resembles Table 1A except that each cell shows the value of p in addition to the set(s) of m winners and the total misrepresentation. The method that we used to select p, based on m and k, is a complex one whose aim is to attain some consistency with respect to average misrepresentation.[5]

The sets of winners in Table 1B are much like those in Table 1A but not identical. As for ties, there is little difficulty. Table 1B shows only three ties, all of them two-way. The lower left-hand corner of Table 1B does not show values of total misrepresentation equal to or even close to zero, as would have been the case had the table been based on $p = 3$ throughout. Thus, lowering p removed the problem of a large number of m-candidate sets that were tied at zero.

Awareness of possible problems with large m and $k = 1$, however, may not be sufficient to deter voters from casting approval votes for "too many" candidates. To reduce the danger that many sets could be tied at zero, one can set $k > 1$. In Table 1B, note that total misrepresentation generally increases as one moves across each row from left to right, in the direction of increasing k, despite the fact that p is also increasing. Setting k to some value greater than 1 pushes total misrepresentation well above zero, thereby removing the danger of many ties and making approval voting more suitable for use with Monroe's PR system.

The larger the electorate, however, the less will be the danger of having many sets tied at zero when $k = 1$. This is because a larger number of voters increases the chance that, for any given set, there will be at least one voter who does not approve of any candidate in the set and who therefore will necessarily be misrepresented. But even though no sets have misrepresentation of zero, there can still be a number of sets whose misrepresentation totals are closely bunched slightly above zero—a result that may be almost as unsatisfactory as having many sets tied at zero.

We also came across a second potential difficulty concerning the use of Monroe's system with approval voting. An example will illustrate. With $k = 1$, suppose that $n = 87$ voters, 15 from party A and 72 from party B, are to choose $m = 3$ winners from $M = 4$ candidates, of whom one is from A and three are from B. Suppose, first, that all 15 A-voters approve of only the single candidate of their party and all 72 B-voters approve of all three of their candidates. These voting strategies, one would think, would be the natural ones. Total misrepresentation, however, is 14 for any set of 3 potential winners that includes the A-candidate, but it is 15 for the set that consists of all three B-candidates. So the A-candidate is a winner even though the strength of party A in the electorate is only about 17%—far below 1/4, the threshold required (Brams 1975, chap. 3) for a party to win a seat in the case of cumulative voting with two parties and 3 winners.

Suppose now that the B-voters somehow split themselves into three groups of 24 voters each, with those in each group voting to approve of a different one of the three B-candidates and no one else. Because total misrepresentation remains at 15 for the set consisting of all three B-candidates but rises to 24 for any set containing the A-candidate, the A-candidate is no longer a winner. Thus the result that is consistent with the cumulative-voting threshold can be achieved, but only if the B-voters adopt a voting strategy that requires coordination among them.[6]

In summary, when Monroe's system is used with approval voting and $k = 1$, there is a potential problem with ties, and another problem with unduly low thresholds for minority candidates.

Some Issues That Are More Technical

Other matters arise in the application of Monroe-type systems generally—not just under approval voting—but they are sufficiently technical that we have relegated them to appendixes. Appendix A deals with easing the computational burdens for large electorates, B with ties, and C with filling of vacancies. We next turn to further major extensions of Monroe's system.

FRACTIONAL ASSIGNMENT

We will speak of *integer assignment* if the x_{ij}'s are forced to be integers, as in (6), and *fractional assignment* if they are allowed to be fractional. Except in a brief footnote, Monroe did not consider fractional assignment of candidates to voters. Thus, instead of assigning each winning candidate to 8.25 voters for $m = 4$ and $k = 1$, he assigned one of them to 9 voters and the other three to 8 voters; and for $m = 2$, the voters were split 17 and 16 between the winners rather than 16.5 being assigned to each winner.

Up to this point we, too, have refrained from considering fractional assignments, largely to maintain consistency with Monroe and make the exposition easier. Now, however, we examine the possibility of assigning exactly kn/m voters to each winning candidate, even when kn/m is not an integer. Under such provision for fractional assignment, each winner always represents exactly the same number of voters, which is not the case for integer assignment.

Mathematically and computationally, fractional assignment is easily handled through mixed-integer programming. The x_i's are still constrained to be integers, but the x_{ij}'s are not. Thus,

$$x_{ij} \leq 1 \quad \text{for each } (i, j) \text{ combination} \tag{9}$$

replaces (6). In addition, (7) is used, with $B = kn/m$, regardless of whether kn/m is an integer, so (3)-(4) are never used. The full set of constraints is therefore (1), (8), (7), (5), and (9).

Although the provision for fractional assignment may seem to be a minor emendation, it can result in significant differences, especially in small electorates. First, fractional and integer assignment can yield different sets of winners. For example, for $m = 5$ and $k = 2$ with voting based on rankings, we found that the winning set (unique in both cases) changed from (D, F, H, I, T) in Table 1A to (F, G, H, I, T) under fractional assignment. Second, if each voter were cloned $(m - 1)$ times, thereby producing an electorate m times as large, one would think that the set of winners ought not to change. This property is always satisfied under fractional assignment but not under integer assignment.[7]

Even with provision for fractional assignment, not many of the x_{ij}'s will actually be fractional. For example, for $m = 4$ and $k = 1$ for either approval voting (with $p = 1$) or voting based on rankings, just three of Monroe's 33 voters had fractional x_{ij}'s. (For both types of voting, our solution gave a 3/4 assignment of one winner to each of these three voters, together with a 1/4 assignment of each of the other three winners to one of those same voters.) In the case where kn/m is an integer, there will always exist a solution that minimizes total misrepresentation and has no fractional x_{ij}'s at all, even though the constraints permit fractional x_{ij}'s.[8]

NONINTEGER *k*

Once provision is made for fractional assignment, it becomes possible to use a noninteger value for k. One's first reaction might be that noninteger k is an unneeded complexity. It may be worth considering for some situations, however.

Consider the following example under approval voting: there are two parties, A and B; $M = M_A + M_B$ candidates, M_A from A and M_B from B; n voters, Pn from A and $(1 - P)n$ from B, all of whom vote for all candidates of their own party and for no one else; and $m = 3$ winners. Suppose first that $k = 1$. As in the example discussed at the end of the section on approval voting and in note 6, A will win a seat even if P is below the cumulative-voting threshold of 1/4, provided that P exceeds 1/6. This leads to the difficulties noted earlier (namely, the B-voters can stop A from winning a seat only by carefully coordinating their votes).

But now let $k = 3/2$ rather than 1, and let all voters continue to approve of all candidates of their own party and no one else. The threshold for party A to win a seat then rises from 1/6 to 1/4, the cumulative-voting threshold.[9] Thus, the threshold problem disappears.

If there are to be $m = 2$ winners rather than 3, the threshold for A to win a seat is $P = 1/4$ if $k = 1$, but it increases to the cumulative-voting threshold of $P = 1/3$ if k is chosen to be 4/3. So one avoids the threshold problem by setting $k = 4/3$ rather than 1.

Unfortunately, the cumulative-voting thresholds cannot be duplicated exactly (through the choice of k) when $m > 3$. This is because there is more than one threshold to contend with when $m > 3$.

For many years, Illinois voters used cumulative voting to elect the lower house of their legislature from 3-member districts. A disadvantage was that parties were uncertain as to how many candidates they should run in a district; they could suffer by running either too many or too few and, in fact, did on occasion (Brams 1975, chap. 3). In our model with $m = 3$ and $k = 3/2$ under approval voting, however, the threshold for party A to win a seat is $P = 1/4$, the cumulative-voting threshold, regardless of what M_A and M_B, the numbers of candidates, are. In this respect, therefore, the system is better than cumulative voting, at least for the case of $m = 3$ (as well as for $m = 2$).

THE CHAMBERLIN-COURANT SYSTEM

As Monroe (1995, 936) noted, Chamberlin and Courant (1983) proposed a PR voting system that is effectively the same as his, except that it (i) does not require "constituency" sizes to be equal or nearly equal and (ii) uses weighted voting in the legislature to make up for variations in constituency size. That is, the Chamberlin-Courant system allows the number of voters assigned to a winning candidate to vary across different winners. In turn, the winners cast these numbers of votes in the legislature.

Solutions for sets of winners under the Chamberlin-Courant system can be found through integer programming. The objective function and the constraints are the same as under Monroe's system, except that constraint (3) is dropped (which is equivalent to retaining it with $L = 0$) and constraint (4) is applied with $U = n$. Although Chamberlin and Courant defined their system only for $k = 1$, it can be used for any k. The constraint set is then composed of (1), (8), (4) with $U = n$, (5), and either (6) or (9). If k is an integer, using (9) rather than (6) does not alter any solutions for winners, but (9) must be used in case k is not an integer. Our findings that deal with grouping of like ballots, with ties, and with filling of vacancies (in Appendixes A, B, and C) are fully applicable, except as noted, under the Chamberlin-Courant system.

If $k = m$, the Monroe and Chamberlin-Courant systems are equivalent, whatever the μ_{ij}'s. This is because their sets of constraints are equivalent and thus lead to the same solutions.

Although Chamberlin and Courant did not mention it, their system, like Monroe's, is applicable not just for voting based on rankings but also for approval voting or, in fact, for any kind of voting from which misrepresentation values, μ_{ij}, can be derived. Under either their system or Monroe's, with any type of voting, a set of winners that minimizes total misrepresentation can, of course, always be extracted from an optimal solution to the integer program. But the Chamberlin-Courant system also requires the determination of the weight that each winner carries in the legislature.

It might appear that the weights could be obtained from the integer-programming results simply by taking $\Sigma_j x_{ij}$, the number of voters to whom candidate i is assigned, as the weight for candidate i. Doing this will work in some cases but not in others. It will always give correct weights if voting is based on rankings and every voter ranks all M candidates, or if the μ_{ij}'s are otherwise such that no two of them are the same for the same voter. In these cases, ties from differing sets of x_{ij}'s for a given set of winners are not possible.

It is observed in Appendix B that ties of this type are inconsequential under Monroe's system, because the x_{ij}'s are not used for anything under his system. Under the Chamberlin-Courant system, by comparison, multiple optimal solutions with different values of the x_{ij}'s can cause a problem. One cannot just use $\Sigma_j x_{ij}$ as the weight for candidate i, because the x_{ij}-values will differ among different solutions. Instead, one must run a simple supplemental calculation once the set of m winners has been obtained from the integer program.

To illustrate how this is done—and the problem corrected—consider first an example with voting based on rankings. Let $m = 3$ and $k = 1$, and suppose that the integer program yields candidates A, B, and C as the winners. For a voter who leaves all three winners unranked, 1/3 of a unit of weight (or possibly zero) could be contributed to each of A, B, and C; otherwise, one unit could accrue to the winner to whom the voter gives the top rank. As a second example, let $m = 4$ and $k = 2$ under approval voting, and let A, B, C, and D be the winners. A voter who approves of all winners except C would contribute 2/3 of a unit to each of A, B, and D; a voter who approves of only C and D would contribute one unit to each; and a voter who approves of only B would contribute one unit to B and 1/3 of a unit (or possibly zero) to each of A, C, and D.

One can conceive of a spectrum of PR voting systems with the Chamberlin-Courant system at one extreme and our fractional-voting system at the other. Between, there will be a continuum of voting systems, based on the degree to which the number of voters assigned to a candidate is forced to be uniform among winners. Mathematically, the continuum can be defined in terms of a variable q that runs from 0 to 1. The constraints are then (1), (8), (3) with $L = qkn/m$, (4) with $U = qkn/m + (1 - q)n$, (5), and (9). The Chamberlin-Courant system has $q = 0$; the fractional-voting system, $q = 1$. Although the Monroe system cannot be placed exactly along this continuum, it falls, roughly speaking, at a q-value close to 1, but not equal to 1 unless *kn/m* is an integer.

The matter of weighted-versus-unweighted voting need not be linked to the value of q. Although Chamberlin and Courant advocated weighted voting for $q = 0$, and Monroe favored unweighted voting for q at or near 1, either option could be used anywhere along the continuum. If unweighted voting is used and q is not close to 1, however, the representation might be far from proportional. At $q = 1$, of course, weighted voting with weights equal to constituency sizes is equivalent to unweighted voting.

TULLOCK'S SYSTEM

Tullock (1967, chap. 10) proposed a PR voting system under which each voter votes for just one candidate, every candidate who receives at least one vote is a winner (so that *all* the candidates are assured of being winners if they vote for themselves), and weighted voting is used in the legislature. The weights carried by the legislators in legislative voting are equal to the numbers of votes cast for them. Candidates who receive just one vote, though they can join the legislature, could be restricted with respect to legislative activities other than voting.

From a mathematical standpoint, Tullock's system can be subsumed within our integer-programming framework as an elementary special case of the Chamberlin-Courant system with $m = M$ and $k = 1$. One can define μ_{ij} to be 0 if candidate i is the top choice of voter j, and 1 otherwise. Although running the integer program will yield the correct solution, there is no need to do so since obviously the solution will just have $x_i = 1$ for each i and $x_{ij} = 1 - \mu_{ij}$ for each (i, j).

Tullock's system has several virtues. It is conceptually simple and easily understood. It is nonmanipulable, because voters have no incentives to vote insincerely for strategic reasons (e.g., to prevent a candidate from being a representative, or to support a second choice rather than a first choice who has little chance of winning). Its representativeness of the electorate is perfect, at least insofar as voters have complete information about candidates' positions and insofar as they vote for themselves in cases where no other candidate mirrors their full views.

On the other hand, two significant objections can be raised against Tullock's system.

First, the size of the legislature may be enormous. Tullock anticipated this objection. He offered some suggestions for operating a large legislature; and some others for holding down its size, such as imposing a maximum on the number of members, and a minimum on the number of votes for a candidate to be elected. But of course these restrictions would make his system vulnerable to strategic voting, undermining its aforementioned nonmanipulability depending on how tight the constraints are.

To hold down the size, one could use some form of the Chamberlin-Courant system with a value of m that is below M but still quite high. Voters would then need to indicate more than just a single first choice; but that would seem desirable even under the pure Tullock system (for which $m = M$), because otherwise the system would have no good way to handle a vacancy.

The second objection is with weighted voting itself, which Tullock's system uses and which we discuss in the next section. The same objection can also be raised, as Monroe has done, against the Chamberlin-Courant system.

WEIGHTED VOTING

We now briefly consider some weighty objections to weighted voting. These objections are perhaps best represented by analyses, such as those based on the Banzhaf and Shapley-Shubik power indexes (Shapley and Shubik 1954; Banzhaf 1965; Riker and Shapley 1968; Brams 1975, chap. 5), that show that legislators with higher weights usually have power greater than what is warranted by their larger constituency sizes. Especially for the weighted-voting PR systems discussed in this article, however, there are factors that may mitigate some of the liabilities of weighted voting:

1. Constituencies are self-selecting under Tullock's system and may be largely so under the Chamberlin-Courant system. The aforementioned analyses of weighted voting have not dealt with self-selecting constituencies or their implications.

2. These constituencies will change with every election, as will the weights. Thus, any adverse effects of weighting need not be long-lasting. In particular, voters will have frequent

opportunities to reduce the weights of heavily weighted legislators.

3. Self-selecting constituencies can be expected to be quite homogeneous. By contrast, a geographical constituency may well be so heterogeneous that its legislator will often cast a vote that is opposed by almost half the constituency; the resulting misrepresentation is aggravated if that legislator also carries a high weight. But a legislator with a homogeneous self-selected constituency, whether the weight is high or low, is not likely to cast a vote that is opposed by more than a small part of the constituency.

4. Arguments that voter power is proportional to the square root of constituency size, which led to the conclusion that disparate constituency sizes should be avoided (Banzhaf 1966, 1968), rest explicitly on the assumption of heterogeneous constituencies.

5. The acceptability of weighted voting may depend partly on whether the legislative body is large or small. As Banzhaf (1965) and Riker and Shapley (1968) themselves indicated, the adverse effects of weighted voting tend to diminish as the size of the legislature increases.

6. The acceptability of weighted voting may vary according to the degree to which a legislative body serves as a vehicle that simply transmits voters' preferences, as opposed to one that also provides a setting for persuasion and bargaining. For example, if the Electoral College in the United States were hypothetically set up with each state's electoral votes proportional to its number of voters, and these electoral votes were split in proportion to the state's popular vote instead of being cast as a bloc for the plurality winner, there would surely be less complaint about weighted voting, because it would mirror the popular vote and thus be transmitting perfectly. But for a legislature where logrolling and vote trading play a heavy role, arguments against weighted voting are telling. These arguments include those involving the nonlinear relationships between voting weight and voting power (for a recent appraisal of voting-power indexes and several related paradoxes, see Felsenthal and Machover 1995).

7. The adverse effects of weighted voting will depend, to some extent, on the values of the weights. These effects may be less if the values can vary over time, enabling shifts in

legislative voting strength to take place.

8. Weighted legislative voting should not, by itself, disqualify a PR system from being acceptable. Other criteria, such as degree of nonmanipulability and degree of representativeness, which we take up next, are also important to consider in comparing different systems.

NONMANIPULABILITY

Our main focus in this article has been on (i) showing the usefulness of integer programming in determining winners under various PR voting systems and on (ii) developing a framework, based on integer programming, that subsumes a number of systems. We do not attempt here to give detailed evaluations of any of the systems with respect to different criteria. However, some comments about two of the most important criteria, nonmanipulability and representativeness, are in order.

Concerning nonmanipulability, Monroe (1995) has already made several observations. He suggested that his system will be relatively resistant to unwanted manipulation if it is used with approval voting and with the μ_{ij}'s determined accordingly. This appears to be a reasonable conjecture in view of different studies that have found approval voting to have favorable nonmanipulability properties.

Monroe also contended that his method, if used with rankings, would be easy to manipulate and therefore impractical. He reasoned that, because the Borda system is highly manipulable, his system, when used with rankings, would be likewise vulnerable since, for the special case of $m = 1$, it is the same as the Borda method.

We have already noted the variety of systems that fall under the broad framework developed in this article. A system that selects the m winner(s) to be the m candidate(s) with the highest Borda score(s), whether $m = 1$ or $m > 1$, is a special case within this framework, with the μ_{ij}'s based on rankings and with $k = m$ and $q = 1$. But another special case with the μ_{ij}'s based on rankings, a system that can be arbitrarily close to Tullock's, arises if one chooses $k = 1$, sets m slightly below M and q at or near 0, and uses weighted voting. Because the Borda system with

$m = 1$ (highly manipulable) and the Tullock system (essentially nonmanipulable) are both subsumed by our framework, it is apparent that PR voting systems based on rankings can run the gamut with respect to manipulability.

REPRESENTATIVENESS

Representativeness has different meanings. Consider the simple case in which all issues before a legislative body have only two sides. Suppose that, on every issue, the majority side in the legislature is the same as the majority side in the electorate. Although a legislature satisfying this property would reflect the desires of the electorate, it would hardly be considered representative if, on every issue, it voted unanimously (or nearly unanimously) on the majority side. This could occur, even with a closely divided electorate, under very different election systems, including ones in which members are elected either by district (if the districts are similar enough in their breakdowns between majority and minority) or at large.

At a more stringent level of representativeness, one could require the percentage split in the legislature to be the same as the percentage split in the electorate on every issue. At least in theory, Tullock's system would satisfy this criterion, because any voters who could not otherwise find a candidate in agreement with all their views could vote for themselves and thus bring about perfect representation.

Whether judged by either criterion—the majority in the legislature on each issue is (i) the majority in the electorate or (ii) a perfect reflection of the majority percentage in the electorate—Monroe's system may be unrepresentative to the extent that the candidate who is assigned to a voter is not the voter's first choice. (Of course, his system is aimed at minimizing such misrepresentation.) In some cases, however, improvement in representativeness can result from nothing more than a change in the field of candidates. For example, if $2n/m$ voters each rank candidate A first and candidate B second but have little enthusiasm for B, then Monroe's system will assign no more than half of these voters to A and assign the remainder to a candidate unappealing to them. Improvement will occur, however, if the ballot at the next election includes not only A but also a clone of A.

The representativeness of some PR systems discussed in this article can run amok of certain anomalies, two of which we illustrate next. First, suppose that $m = 5$ winners are to be chosen from $M = 6$ candidates, three (A1, A2, and A3) from party A and three (B1, B2, and B3) from party B, with the μ_{ij}'s based on approval voting and with $q = 1$. Suppose that 3/5 of the n voters approve of just A1, A2, and A3, and that the other 2/5 approve of only B1, B2, and B3. There are essentially two possible outcomes, based on whether one of the A-candidates or one of the B-candidates is the loser. One would expect a PR system to choose the three A-candidates and two of the B-candidates as the winning set. This is what happens for $k = 1$, 2, and 5 candidates per voter. But for $k = 4$ the two outcomes are tied, because total misrepresentation is $(8/5)n$, regardless of whether the winning set excludes an A-candidate or a B-candidate. Worse yet, any winning set for $k = 3$ has only two of the A-candidates, because total misrepresentation is $(3/5)n$ if an A-candidate is excluded and $(4/5)n$ if a B-candidate is excluded.[10]

As a second example, consider an election in which there are two issues and $n = 60$ voters. Suppose that the voter positions on the issues are (Y, N), or (Yes, No), for 27 voters; (N, Y) for another 27; and (Y, Y) for the remaining 6. There are $M = 4$ candidates, A, B, C, and D, with respective positions (Y, Y), (Y, N), (N, Y), and (N, N). With the μ_{ij}'s based on rankings, let Monroe's system ($q = 1$, $k = 1$) be used to choose $m = 3$ winners. Suppose that all voters vote their sincere preferences, and that these are (B, D, A, C) for the first 27 voters, (C, D, A, B) for the second 27, (A, B, C, D) for 3 of the last 6, and (A, C, B, D) for the remaining 3. Total misrepresentation is then 26 for the set (B, C, D) (for which candidate D is assigned to 10 of the first 27 and 10 of the second 27 voters) and 28 for the set (A, B, C). Thus (B, C, D) wins; the legislative body opposes both issues by 2 to 1 even though the electorate favors both by 33 to 27; and candidate A, who would be the electee of the "voice of reason" (Monroe 1995, 935) in a single-winner election, is the sole loser. Because neither A nor any of the other three candidates is a Condorcet winner, these perplexing results are perhaps not so surprising.

HIERARCHICAL PR

So far we have considered a plethora of PR systems that minimize the overall misrepresentation of voters. Integer programming, which is applicable both to ranking systems (e.g., the Borda count) and to nonranking systems (e.g., approval voting), provides a systematic way of finding the set of winning candidates.

In the United States, there are three levels of federal office, with varying numbers filling each level: 1 president and 1 vice president; 100 senators; and 435 representatives. These three levels, written into the federal constitution, are mirrored in all states except Nebraska, which has, like the other states, one governor, but a unicameral rather than a bicameral legislature.

We think it impractical as well as unwise to attempt to change the U.S. Constitution, or state constitutions, to elect senators and representatives nationwide, or state legislators statewide, using the kind of PR system we have proposed (although PR applications to elect legislators from multi-member districts may nonetheless be feasible). Our type of system, we believe, is more appropriate for electing city councils or other representative bodies of, say, 5 to 25 members. Such a PR system could also be adapted, as we will next show, to allow for the election of not only regular members of a council but also a council chair or a mayor.

If separate elections are held for a mayor and a city council, as is usually the case today, the more ambitious candidates, who prefer to be mayor but think it safer to run for city council, are forced to make a difficult choice of which office to run for. By contrast, under a system we call *hierarchical PR*, different levels of representation can be achieved in just one election.

To illustrate how hierarchical PR would work, consider a municipality in which 1 mayor and 5 members of a city council are to be elected. Under hierarchical PR, the mayor would be the one candidate, among all the others running for the city council, who minimizes misrepresentation across all the voters. If voters can rank the candidates, this person would be the Borda winner; if the voters can approve of as many candidates as they like, this person would be the approval-voting winner.

Once chosen, the mayor is eliminated from further consideration. The 5-member council is then elected from the remaining candidates to minimize total misrepresentation.[11]

Suppose, for example, that approval voting is used with $k = 1$ (i.e., one candidate is assigned to each voter) and $q = 1$ (i.e., each candidate is assigned to the same number of voters), and that there are 20 voters who approve of the candidates as follows:

•voters 1-3, 5-7, 9-11, 13-15, and 17-19 (75%) approve of candidate A;
•voters 1-4 (20%) approve of candidate B;
•voters 5-8 (20%) approve of candidate C;
•voters 9-12 (20%) approve of candidate D;
•voters 13-16 (20%) approve of candidate E;
•voters 17-20 (20%) approve of candidate F.

In the race for mayor, A would be elected. In the race for council, (B, C, D, E, F) would be elected (which would also be true if there were other candidates who are approved by fewer voters). In the case of the council, observe that every one of the 20 voters is represented by exactly one of the five winners in such a way that total misrepresentation is zero.

Now consider a situation in which there is no election for mayor but just an election for the 5-member city council (assume the council chooses one of its elected members as mayor). Then the same council as before, (B, C, D, E, F), would be chosen. This is fine, as far as it goes: all the voters, as before, have exactly one representative of whom they approve. The rub is that the most popular candidate, A, loses, because any 5-member council that includes A would misrepresent one voter.[12]

Admittedly, our example is an artificial one designed to illustrate a certain type of possible outcome under hierarchical PR for two offices, one with 1 winner and the other with 5 winners in this case, when $k = 1$. Less artificial is Monroe's example, in which the *Independent* (I) would have been chosen for the first "office," because it is the overall most popular newspaper (recall that misrepresentation in his example was based on ranks).

If there were 2 winners for the next "office" (instead of the 5 we assumed in our city-council example), they would not, of course, include the *Independent*, because it is eliminated.

As we showed in Table 1A, the *Independent* and the *Times* (T) are the two newspapers that would have been chosen if there were two winners for a *single* office. This illustrates the fact that hierarchical PR need not give the same results for a second-level office as does a nonhierarchical election with the same number of winners, and the fact that the winners in the latter election need not exclude the winner of the top office under hierarchical PR.

If there are two or more levels of representation, hierarchical PR has the advantage of not requiring different elections for the different levels (offices), relieving candidates of the necessity of making the kind of difficult choice alluded to earlier (to run for mayor or for city council). Instead, the system itself chooses the most popular candidate for the "highest" office, and the most representative choices for lower offices, making endogenous who is most appropriate at each level.

It is worth noting that a higher office might well have several winners. For example, this office might be a 5-member executive committee, which has administrative functions that it would be unwieldy for, say, a 25-member council, of which the executive committee is a part, to carry out. Although traditionally it is the council or the council chair that selects such a committee, under hierarchical PR it would be the voters who choose this committee.

The remaining council members would be the 20 other candidates who, after the 5 executive-committee members are chosen, minimize—at a second level—the total misrepresentation of the voters. Integer programming would be applied twice, first to determine the composition of the executive committee and then to select the rest of the council, even though only one election to the 25-member council would have to be conducted.

We assume that the voters and the candidates would know that the executive committee would be chosen first. In a campaign, therefore, those "running" for the executive committee would presumably make their appeals sufficiently broad so that, together with a group of 4 other candidates, they might "cover" the electorate. By comparison, those running for the regular council would presumably make narrower, more ideological appeals, seeking only enough support so that, together with 19 other council candidates, they also "cover" the electorate.

Other than at-large elections in which the top vote-getter assumes a top position in a governing body, we know of no other voting system in which hierarchical levels can be derived from just the ballot information in one election. We hesitate, however, to offer prescriptions about the optimal number of levels—except to observe that most political systems have a single top office-holder (executive committees notwithstanding). Below the president, governor, or mayor, there is usually a bicameral legislature or a unicameral council.

Practically speaking, we think hierarchical PR is most relevant to the election of single executives or small executive committees, on the one hand, together with larger councils or boards of which they are a part, on the other. Municipal and county elections in the public sphere, and corporate boards and their executive committees in the private sphere, come quickly to mind. In academia, of course, there is hierarchization in both departments and faculties, to which we think hierarchical PR would also be applicable.

Hierarchical PR can be used with most of the PR voting systems considered in this article. One exception is the Tullock system, which would not be suited to electing a specified number of persons (one or more) to the top office. Additionally, in cases where there is to be only one winner at the top level, the use of a system based on rankings would be open to question, because the highly manipulable Borda system would be used for the top position.

CONCLUDING REMARKS

In this article we have shown how integer programming provides a means for determining election winners under different PR voting systems. Extending the stimulating work of Monroe (1995), we have also developed a framework, based on integer programming, that encompasses a large class of PR systems. Within this framework, the variables that a designer of a PR system can set include the nature of the misrepresentation values (à la Monroe), the number of candidates assigned to a voter (k), a variable (q) that controls the extent to which the number of voters assigned to a candidate can vary among winners, and whether or not weighted legislative voting is to be used. Other variables are the number of candidates (M), the number of winners (m), and the number of voters (n).

Which PR system is best? Does Tullock's system deserve resuscitation after years of neglect? Is the Chamberlin-Courant system viable in some form? Or are both these systems fatally flawed because of their use of weighted voting? If so, should one turn to one of the Monroe systems, perhaps one with approval voting, or maybe to some hybrid system within the framework? Or is there a method outside of our framework, such as the Hare system, that is better than anything within the framework?

The PR systems of this article involve a number of complex issues that are difficult to summarize. To give some structure to them, however, we offer the following thoughts and recommendations:

1. The best PR system to use will depend on conditions such as the nature of the institution, the size of the electorate, its diversity, and the number of parties (if any).

2. Most PR systems we considered have pitfalls not obvious on first examination. Some flaws may be as yet undetected in these systems; this will require further scrutiny.

3. Monroe, we believe, was a bit too pessimistic about PR systems based on rankings and a bit too optimistic about those based on approval voting. Conditions under which each balloting method works best deserve further study.

4. It is not apparent what the value of k, the number of winning candidates assigned to each voter, should be. The answer may depend not only on "objective" circumstances but also on taste. On the one hand, the choice $k = 1$ offers the advantages of simplicity and interpretability, encourages voters to support fewer candidates in the case of approval voting, and, in concept at least, gives each voter a single champion of his or her views. On the other hand, we showed cases where $k = 1$ gave less representative results than a larger value k.

5. Fractional assignment has several advantages over integer assignment, and no apparent disadvantages. It is, however, less intuitive.

6. At least for a two-party system, the PR method that elects legislators in 3-member districts using approval voting and $k = 3/2$ appears promising. The main barrier to adoption may be the difficulty of explaining the use of 3/2 for k.

7. In the context of our PR framework, weighted legislative voting gives a mixed picture. On the positive side, some of the standard arguments against weighted voting do not apply to a Tullock-like system, which scores high on nonmanipulability and representativeness. On the negative side, there are institutional problems, including (i) the apparent nonegalitarian nature of weighted voting, which makes some representatives more equal than others; and (ii) the fact that the time available to weightier representatives to take care of the burden of representation remains finite and unexpandable, even if they are given larger staffs to serve their more numerous supporters.

8. Most PR systems can be extended to provide hierarchical levels of representation, whereby the most representative candidates for *different* offices can all be chosen in one election rather than being required to run in different elections. The idea of hierarchical PR seems sufficiently promising that it deserves to be explored further. It strikes us as not only an efficient approach to building representation at different levels—because it requires only one election—but also more democratic than having, say, an elected body choose a leader, as occurs today in most parliamentary systems.

To conclude on this last point, we think it is the voters themselves who should have the opportunity to choose, for highest office, the most representative candidate or candidates (e.g., a prime minister and also possibly a cabinet).[13] This process can then be repeated at lower levels, with voters in each case choosing representatives who minimize their misrepresentation at those levels. Thereby voters would better be able, with hierarchical PR, to express themselves at *all* levels, which seems to us the cornerstone of democratic representation, especially in a federal system.

APPENDIX A: LARGE ELECTORATES

Integer programming is computationally demanding. At some point a PR integer program will become so large that computing time will be prohibitive. In order to render the technique more suitable for larger electorates, it is useful to examine possible ways to cut computer time. We recognize, however, that the computational burden is likely to become less of an issue as rapid improvements in computer technology continue.

The following modification does not affect the solution for the winners, but it reduces computer time because it cuts both the number of x-variables and the number of constraints. In a large electorate, there will be cases where two or more voters mark their ballots alike and thus have identical misrepresentation values. The modification consists of consolidating, into one group, each set of voters who mark their ballots the same way for a given race.

We redefine the index j so that it applies to these groups rather than to individual voters. Although j runs, as before, from 1 to n^*, n^* becomes the number of groups (of voters who all vote alike) and is no longer equal to n, the number of voters.

Let v_j denote the number of voters in group j ($j = 1, 2, ..., n^*$). We continue to use n to denote the number of voters, where now $n = \Sigma_j v_j$. Although x_i has the same meaning as before, x_{ij} is now the number of voters, out of the v_j voters in group j, to whom candidate i is assigned.

The objective function, z, has the same formula as before, but two of the constraints change. Constraint (8), which is the generalization of (2), changes to

$$\Sigma_i x_{ij} = kv_j \quad \text{for each } j, \tag{A-1}$$

which specifies that the number of candidate assignments to the v_j voters in group j is to be kv_j. Constraint (6) becomes

$$x_{ij} \text{ is an integer less than or equal to } v_j, \text{ for each } (i, j) \text{ combination.} \tag{A-2}$$

Thus the constraints are now (1), (A-1), (3)-(4) or (7), (5), and (A-2). L, U, and B are as before, as described after (8).[14]

It is instructive to apply this modification to Monroe's data. Voters 1 and 30 both ranked the candidates in the order (F, T, I), with the other three candidates unranked, and voters 3 and 25 both voted (I, T, G). There were no other cases where voters voted alike. Thus, $v_j = 2$ for the two j-values corresponding to these two groups, $v_j = 1$ for the remaining 29 j-values, $n = 33$ as before, and $n^* = 31$.

To illustrate our solution for $m = 4$ and $k = 1$, (F, G, I, T) is the winning set of candidates with total misrepresentation of 13, just as before. The new solution also yields $x_{2j} = 2$ for the first group, $x_{5j} = x_{6j} = 1$ for the second group, and 0 for the remaining x_{ij}'s in those two groups. Hence, F is assigned to both voter 1 and voter 30. Whether T is assigned to voter 3 and I to voter 25, or vice versa, does not matter since the effect on total misrepresentation is the same.

The type of voting system that is used affects n^*—which, in turn, affects computing requirements since a lower n^* should reduce computer time. For approval voting, n^* cannot exceed $2^M - 2$ (disregarding ballots that rate all candidates the same). For voting based on rankings, the maximum n^* is much larger, except when M is very small, since it is equal to $M!$ even if each voter ranks all M candidates. If one allows for voters to leave candidates unranked (but only at the bottom of the rankings), the maximum n^* is[15]

$$M! + M!/2! + M!/3! + \ldots + M!/(M-2)! + M!/(M-1)! \,. \qquad \text{(A-3)}$$

The value (A-3) is bounded above by $(e - 1)M!$, or about $1.72M!$

For voting based on rankings, the maximum n^* would be $M!$, instead of about $1.72M!$, if a ballot with g unranked candidates at the bottom ($g = 2, 3, \ldots, M - 1$) were tallied as $(1/g!)$-th of a ballot for each of the $g!$ ways of completing the ballot. Even though such a change would decrease the maximum n^*, however, the actual n^* could be either higher or lower since the splits into $g!$ parts affect n^* both positively and negatively. It is possible for the change to lead to a difference in the solution for the winners,[16] except that for $k = m$ there can be no difference because in that case the winners are based on Borda scores, which will not change.

Fractional assignment, discussed earlier, can be used together with the technique for grouping of like ballots that is described in this appendix. Let (A-2*) denote the relation (A-2) with the words "an integer" deleted. The applicable set of constraints with fractional assignment is then (1), (A-1), (7) with $B = kn/m$ (which need not be an integer), (5), and (A-2*).

Because fractional assignment and integer assignment are more likely to produce different sets of winners in smaller electorates, the earlier discussion was mainly in the context of such electorates. For larger electorates, however, fractional assignment also deserves serious consideration, for a different reason: it could bring about savings in computer time because it sharply reduces the number of x-variables constrained to be integers. (If k is not an integer, provision for fractional assignment would be mandatory rather than optional.)

APPENDIX B: TIES

The minimum value of total misrepresentation may be attainable at more than one set of values of the x-variables. Commonly, in fact, the minimum is attainable at differing sets of values of the x_{ij}'s. This type of multiple solution (or tie) causes no problem under Monroe's system, however, because the x_{ij}'s reflect assignments for computational purposes only. Indeed, voters will generally not even be aware of what candidates were assigned to them by the solution.

A second type of tie, though less frequent (especially in large electorates), is more significant when it does occur. If minimum total misrepresentation is attainable at more than one set of values of the x_i's, then there will be a tie with respect to who the winners are. Although integer-programming software will find a solution that minimizes total misrepresentation, it will not then go on automatically to find other solutions that also attain that minimum, but with a different set of winners. One can check for ties by running additional integer program(s), however. In fact, we searched for ties in all 30 elections shown in Tables 1A and 1B and found that there were no ties other than those indicated in the tables. One checking method consists of running an integer program that is the same as the one run initially, except that there is an added restriction that rules out the set of winners that was already found (but rules out nothing else extra).

To illustrate in the case of $m = 4$ and $k = 3$ in either Table 1A or Table 1B, our initial solution showed (F, G, I, T) to be the winning set. To stop a rerun from yielding this set again, we added the restriction $x_1 + x_4 \geq 1$, which forces either D or H into the set of winners. Because the new solution had greater total misrepresentation than the original one, we could conclude that there was no tie. Had we found a tie, we would have had to run another integer program to check for a further tie.

To break a tie, one could choose randomly, with equal probability, from among the tied sets. Other tie-breaking techniques might also be devised.

APPENDIX C: FILLING VACANCIES

Any voting system has to be able to deal with a vacancy that arises before the end of a legislator's term. Moreover, the method of filling the vacancy must be such that no legislator except the vacating member will lose a seat.

With any of our PR integer-programming models, it is easy to fill a vacancy based on the original ballots and do so with no danger that an earlier winner will become a loser. For example, in either Table 1A or Table 1B, consider the case of $m = 3$ and $k = 1$. F, I, and T are the winners. Suppose that T vacates, and that D, G, or H is to be chosen as a successor, based on the original ballots. After redoing the ballots with T removed,[17] one simply runs the usual integer program, except that two extra constraints, $x_2 = 1$ and $x_5 = 1$, are added. Without such added constraints, an earlier winner could be unseated.[18]

Under the Chamberlin-Courant system, although weights will be determined anew, all nonvacaters will still be able to retain their seats when a vacancy is filled. A change in weight could be either up or down.[19]

Notes

1. Such resistance might be reduced, and the support for PR systems increased, as a consequence of recent Supreme Court decisions that, by disallowing some of the more bizarre-looking districts in several Southern states, may bring about diminished minority representation.

2. The only previous voting application of integer programming that we are aware of pertains to constrained approval voting; see the comment by Potthoff (1990) on the article by Brams (1990), and see Straszak et al. (1993). Elsewhere in political science, Herlihy (1981) applied mixed-integer programming to a districting problem in the Republic of Ireland. In connection with the problem of apportioning representatives to the U. S. House of Representatives and similar bodies (Balinski and Young 1982; Ernst 1994), integer programming apparently has not been applied but certainly could be used.

3. We give any unranked candidate the average of the voter's unused ranks. For example, if a voter ranks only his or her two top choices out of $M = 6$ candidates, the ranks are 1, 2, 4.5, 4.5, 4.5, 4.5.

4. Comparison of the results in the first column of our Table 1A with those of Monroe (1995, Table 2) reveals disagreements between them over both total misrepresentation and winning candidates. Our rule for assigning ranks to unranked candidates was the same reasonable rule that Monroe (1995, 928) said he used: each unranked candidate was assigned the average of the unused ranks. We suspect, though, that Monroe, in his calculations, did not use this rule; rather, it appears that if a voter left, for example, 4 of the 6 candidates unranked, then Monroe's ranks were (1, 2, 3, 3, 3, 3) instead of (1, 2, 4.5, 4,5, 4.5, 4.5). Because such a discrepancy would not in itself prove Monroe's solution method (which was not fully explained) to be faulty, we reran our integer programs using what appeared to be the rule that Monroe actually used. This time our results agreed with Monroe's for $m = 1, 2$, and 3, but not for $m = 4$. For $m = 4$ with the altered rule, we found the (unique) set of winners to be (F, H, I, T) compared

with Monroe's answer of (D, F, H, I) tied with (F, G, I, T); total misrepresentation was 12 compared with 13.

5. Briefly, for each (m, k), we chose $p = p(m, k)$ to be the value of p (1, 2, 3, 4, or 5) for which $E(m, k, p)$ was closest to $E_0(m, k)$ (using the higher value of p in case of a tie), where $E(m, k, p)$ denotes the average misrepresentation per voter for given (m, k, p), and $E_0(m, k)$ denotes the "ideal" misrepresentation. We chose $E_0(m, k) = E_0(k) = (k + 1)/4$, which is a 50-50 mixture of $E_{01}(k) = 1/2$ and $E_{02}(k) = k/2$. Assume, for simplicity, that the p candidates whom a voter approves of are a random draw from the $M = 6$ candidates on the ballot, and let r (a random variable) denote how many of these p candidates are winners. Define $C(N, R) = N!/[R!(N - R)!]$ if $0 \leq R \leq N$, or 0 otherwise. The average misrepresentation is then

$$E(m, k, p) = \sum_{r=0}^{k} (k - r)\, P(m, p; r),$$

where $P(m, p; r) = C(6 - m, p - r)\, C(m, r) \,/\, C(6, p)$ is the probability that exactly r of the p approved candidates are winners. Thus, for $m = 5$ and $k = 3$, $E(5, 3, p)$ is 13/6, 8/6, 3/6, 0, 0 for $p = 1, 2, 3, 4, 5$. Consequently, $p(5, 3) = 2$, since $E(5, 3, 2) = 8/6$ is closer to $E_0(3) = 1$ than $E(5, 3, 3) = 3/6$ is.

6. For the general case of approval voting under Monroe's system with $k = 1$, two parties, m winners, one A-candidate, and m B-candidates, party A will win one seat if its strength exceeds the threshold of $1/(2m)$ of the electorate (equal to 1/6 for $m = 3$) when all A-voters vote just for the A-candidate and all B-voters approve of all m of the B-candidates. If the B-voters coordinate their strategies so that only $1/m$ of them vote for each of their m candidates, then the threshold for A to win one seat becomes $1/(m + 1)$, the same as the cumulative-voting threshold.

7. Mathematically, the minimization problem is virtually the same after the cloning as it is before if there is fractional assignment, but not if there is integer assignment. In the example earlier in the paragraph, if Monroe's 33 voters are each cloned four times to produce an electorate of size $n = 165$, then (i) kn/m becomes an integer, 66; (ii) the winning set is

(F, G, H, I, T) not only under fractional assignment but also under integer assignment; and (iii) total misrepresentation is 305 under both types of assignment, as compared with 61 under fractional assignment and 59.5 under integer assignment before the cloning.

8. We do not give a detailed proof of this last statement but note that the proof is based on two main points. First, if the x_i's are held fixed, so as to correspond to any specific potential set of m winners, then the minimization problem for the specific set—with no integer restrictions on the x_{ij}'s—is a special case of what is known as the *capacitated transportation problem.* Second, there exists an all-integer optimal solution to a capacitated transportation problem provided that all the constraint constants [B and the right-hand sides of (8) and (9), in the present case] are integers. See Dantzig (1963, 377-380).

9. For any permissible values of P, m, and k, the minimized total misrepresentation, after division by n, is $(k - 1)P + (k/m - P)$ if A receives one seat (provided that $P \leq k/m$), and it is kP if A receives no seats. Equating these two expressions and solving for P yields $P = k/(2m)$, the threshold for A to win one seat. If one equates this expression for P with $P = 1/(m + 1)$, the cumulative-voting threshold, and then solves for k, the result is $k = 2m/(m + 1)$, which is the value of k required to duplicate the cumulative-voting threshold. The value is $k = 3/2$ for $m = 3$ winners and $k = 4/3$ for $m = 2$.

10. The easiest way to verify informally the disconcerting result for $k = 3$ is to let $n = 5$ and construct two 5×5 grids in which the rows of each grid are for the 5 voters (3 A-voters and 2 B-voters), the columns of the first grid represent the candidate set (A1, A2, B1, B2, B3), and the columns of the second are for the set (A1, A2, A3, B1, B2). Into the 25 squares of each grid one then places 15 X's (where each X represents the assignment of a voter to a candidate) so that there are 3 X's in each row and each column, and so that the total misrepresentation is as low as possible. One finds that the minimal misrepresentation is 3 for the first grid and 4 for the second. The result for $k = 4$ can be demonstrated similarly.

11. If the μ_{ij}'s (the misrepresentation values) are based on rankings of the candidates,

there are two ways to define these values after elimination of the mayor: they can be left as they are, or they can be refigured based on rerankings without the mayor. In the latter case, μ_{ij} stays the same, decreases by 1, or decreases by 1/2 according to whether, respectively, voter j ranks candidate i above, below, or tied with the mayor. It is not clear which of the two ways is preferable.

12. It is worth mentioning that whoever is chosen as mayor would be supported by only 4 voters and can, therefore, hardly be considered representative of the entire electorate.

13. This is also possible under "coalition voting" (Brams and Fishburn 1992), one aspect of which is "yes-no voting" (Brams and Fishburn 1993).

14. To prove in general that replacing constraint sets (8) and (6) with (A-1) and (A-2) does not affect the solution for the winners, one has to show that, for any set of x_{ij}'s satisfying (A-1) and (A-2) for a given group j, it is possible to distribute the kv_j assignments so that k candidates are assigned to each of the v_j voters and so that (for each i) candidate i is assigned to x_{ij} of these voters. To show this, one can simply cycle the assignments through the v_j voters, one winning candidate at a time. For example, if $v_j = 3$, $k = 2$, and $m = 4$, and if winning candidates A, B, C, and D have respective x_{ij}'s of 2, 0, 3, and 1, then one can assign A to voter 1, A to 2, C to 3, C to 1, C to 2, and D to 3. With some simple alteration, this technique works even if fractional assignments (covered later in Appendix A, and also in the main part of the article) are allowed.

15. The first term in (A-3) is the number of ways of ranking all M candidates. The g-th term ($g = 2, 3, ..., M - 1$) in (A-3) is the number of ways of ranking the candidates if g candidates are unranked at the bottom. (There is, of course, no term for a single unranked candidate at the bottom.)

16. The following example proves that splitting an incomplete ballot into $g!$ equal parts can lead to a solution whose set of winners differs from that obtained when unranked candidates simply receive the average of the unused ranks. Suppose that $n = 60$, $k = 1$, $m = 2$, and $M = 3$,

with the candidates labeled A, B, and C. Let there be 18 voters who vote (C, B, A), 20 who vote (B, C, A), and 22 who place A first but do not rank the other two candidates. With the ballot splitting method, $(\mu_{1j}, \mu_{2j}, \mu_{3j})$ will be (2, 1, 0) for 18 ballots, (2, 0, 1) for 20 ballots, (0, 1, 2) for 11 ballots, and (0, 2, 1) for 11 ballots. Total misrepresentation, when minimized, is 26 for set (A, B), 28 for (A, C), and 23 for (B, C), so (B, C) is the winning set. With the average-rank method, misrepresentation values are as before, except that they are (0, 1.5, 1.5) for each of the last 22 voters. Although total misrepresentation is again 26 for (A, B) and 28 for (A, C), it is now 33 for (B, C), so (A, B) wins.

17. If the μ_{ij}'s are based on rankings, there are two ways of revising them with one candidate removed, just as explained in note 11. The example in note 18 is constructed so that the outcome is the same regardless of which way is used.

18. That this is so is shown by an election under the Monroe system with $M = 4$, $m = 2$, and $k = 1$, where the rankings (D, A, B, C), (A, B, D, C), (C, B, D, A), and (C, D, B, A) each appear on 1/4 of the ballots. The winners are A and C. Suppose that C vacates. A recalculation with C removed yields B and D as the new pair of winners if no constraint is added.

19. Although an *ad hoc* method could be devised to prevent any weights from dropping, this restriction could give rise to new problems, thereby replacing one imperfection with others.

References

Baker v. *Carr*. 1962. 369 U.S. 186.

Balinski, Michel L., and H. Peyton Young. 1982. *Fair Representation: Meeting the Ideal of One Man, One Vote*. New Haven: Yale University Press.

Banzhaf, John F., III. 1965. "Weighted Voting Doesn't Work: A Mathematical Analysis." *Rutgers Law Review* 19 (Winter):317-43.

Banzhaf, John F., III. 1966. "Multi-Member Electoral Districts—Do They Violate 'One Man, One Vote' Principle." *Yale Law Journal* 75 (July):1309-38.

Banzhaf, John F., III. 1968. "One Man, 3,312 Votes: A Mathematical Analysis of the Electoral College." *Villanova Law Review* 13 (Winter):304-32.

Brams, Steven J. 1975. *Game Theory and Politics*. New York: Free Press.

Brams, Steven J. 1990. "Constrained Approval Voting: A Voting System to Elect a Governing Board." *Interfaces* 20 (September-October):67-79.

Brams, Steven J., and Peter C. Fishburn. 1983. *Approval Voting*. Boston: Birkhäuser.

Brams, Steven J., and Peter C. Fishburn. 1992. "Coalition Voting." *Mathematical and Computer Modelling (Formal Theories of Politics II: Mathematical Modelling in Political Science*, special issue edited by Paul E. Johnson) 16 (August/September):15-26.

Brams, Steven J., and Peter C. Fishburn. 1993. "Yes-No Voting." *Social Choice and Welfare* 10 (January):35-50.

Chamberlin, John R., and Paul N. Courant. 1983. "Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule." *American Political Science Review* 77 (September):718-33.

Dantzig, George B. 1963. *Linear Programming and Extensions*. Princeton: Princeton University Press.

Davis, K. Roscoe, and Patrick G. McKeown. 1984. *Quantitative Models for Management*. 2d ed. Boston: Kent.

Ernst, Lawrence R. 1994. "Apportionment Methods of the House of Representatives and the Court Challenges." *Management Science* 40 (October):1207-27.

Felsenthal, Dan S., and Moshé Machover. 1995. "Postulates and Paradoxes of Relative Voting Power—A Critical Re-Appraisal." *Theory and Decision* 38 (March):195-229.

Gould, F. J., G. D. Eppen, and C. P. Schmidt. 1993. *Introductory Management Science*. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Herlihy, Patrick. 1981. "Does Operations Research Have a Role in the Revision of Irish Parliamentary Constituencies?" In *Operational Research '81*, ed. J. P. Brans. Amsterdam: North-Holland.

Monroe, Burt L. 1995. "Fully Proportional Representation." *American Political Science Review* 89 (December):925-40.

Potthoff, Richard F. 1990. "Comment: Use of Integer Programming for Constrained Approval Voting." *Interfaces* 20 (September-October):79-80.

Riker, William H., and Lloyd S. Shapley. 1968. "Weighted Voting: A Mathematical Analysis for Instrumental Judgments." In *Representation*, ed. J. Roland Pennock and John W. Chapman. New York: Atherton.

Shapley, L. S., and Martin Shubik. 1954. "A Method for Evaluating the Distribution of Power in a Committee System." *American Political Science Review* 48 (September):787-92.

Straszak, Andrzej, Marek Libura, Jaroslaw Sikorski, and Dariusz Wagner. 1993. "Computer-Assisted Constrained Approval Voting." *Group Decision and Negotiation* 2 (November):375-85.

Tullock, Gordon. 1967. *Toward a Mathematics of Politics*. Ann Arbor: University of Michigan Press.

TABLE 1A

Integer Programming Solutions for Winners and Total Misrepresentation, for *m* Winners with *k* Candidates Assigned to Each Voter, Using Monroe's Data with Misrepresentation Values Based on Ranks*

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$m = 1$	I; 48				
$m = 2$	I,T; 27	I,T; 114.5			
$m = 3$	F,I,T; 16	F,I,T; 84.5	G,I,T; 198		
$m = 4$	F,G,I,T; 13	F,G,I,T; 70	F,G,I,T; 161	F,G,I,T; 290	
$m = 5$	D,F,H,I,T and F,G,H,I,T (tie); 9	D,F,H,I,T; 59.5	F,G,H,I,T; 142	F,G,H,I,T; 250	F,G,H,I,T; 391

*The entry in a cell for any *m* and *k* shows the set(s) of winners, followed by the value of total misrepresentation.

TABLE 1B

Integer Programming Solutions for Winners and Total Misrepresentation, for *m* Winners with *k* Candidates Assigned to Each Voter, Using Monroe's Data with Misrepresentation Values Based on Approval Voting*

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$m = 1$	I; 5 (p=3)				
$m = 2$	F,I and I,T (tie); 7 (p=2)	I,T; 13 (p=4)			
$m = 3$	F,I,T; 9 (p=1)	G,I,T; 8 (p=3)	G,I,T; 23 (p=4)		
$m = 4$	F,H,I,T; 7 (p=1)	D,F,I,T and F,H,I,T (tie); 15 (p=2)	F,G,I,T; 26 (p=3)	F,G,I,T and G,H,I,T (tie); 43 (p=4)	
$m = 5$	F,G,H,I,T; 7 (p=1)	D,F,H,I,T; 12 (p=2)	D,F,H,I,T; 40 (p=2)	F,G,H,I,T; 47 (p=3)	F,G,H,I,T; 63 (p=4)

*The entry in a cell for any *m* and *k* shows the set(s) of winners, followed by the value of total misrepresentation and the value of *p* that was used. An approval vote was counted for the *p* top-ranked candidates on each ballot, or for all the ranked candidates if fewer than *p* were ranked.